Knowledge Checks for Chapters 06 – 09

DS598 B1 – DL4DS

Spring, 2024

We also suggest you attempt and review the exercises with solutions provided on the UDL website, to help strengthen your understanding. No need to turn those in however.

The LaTeX format of this assignment is available on the course website.

Chapter 06

Problem 6.10 Show that the momentum term $\mathbf{m}_t$ (equation 6.11) is an infinite weighted sum of the gradients at the previous iterations and derive an expression for the coefficients (weights) of that sum.

Hint: Write the equation for $\mathbf{m}_1$ in terms of $\mathbf{m}_0$ and then $\mathbf{m}_2$ in terms of $\mathbf{m}_1$. Then substitute $\mathbf{m}_1$ for the equation for $\mathbf{m}_1$ and so on. Look for a pattern that you can express compactly.

Chapter 07

Problem 7.3 What size are each of the terms in equation 7.19? Express you answer for each term as $\mathbb{R}^{D_i \times D_j}$, where you replace i and j with the appropriate numbers.

Chapter 08

Problem 8.2 What values should we choose for the three weights and biases in the first layer of the model in figure 8.4a so that the responses at the hidden units are as depicted in figures 8.4b-d?

Chapter 09

Problem 9.1 Consider a model where the prior distribution over the parameters is a normal distribution with mean zero and variance σ_ϕ^2 so that

$$Pr(\phi) = \prod_{j=1}^{J} Norm_{\phi_j}[0, \sigma_\phi^2].$$

where j indexes the model parameters. When we apply a prior, we maximize

$$\prod_{i=1}^{I} Pr(\mathbf{y}_i|\mathbf{x}_i, \phi)Pr(\phi).$$

Show that the associated loss function of this model is equivalent to L_2 regularization.

Hint: Calculate the negative log likelihood of the above function and then substitute the equation for the prior distribution and simplify.